

**Goal:** To become acquainted with using python to parse data files and perform more complex data analysis including comparing distribution and performing counting process

**Approach:** The python code was written by writing a class that would contain the parsed data. The data file to parse is passed through the program parameters followed by the name of a file to output the chart to. An example of the usage is as follows:

```
$ ./Coronado.HW.2.py < data_file> < plots utput_output_file >
```

The libraries used for this program where matplotlib, statistics, and scipy.stats. These where used for plotting the data and performing basic statistical operations on the data. After, parsing the individual packets the data was then broken down into two separate data sets for incoming and outgoing packages. Lists for inter-arrival times, elapse times and counts per time interval were created each data set. Basic statistics analysis was performed on the inter\_arrival times and their probability distribution was plotted. The probability distribution of counts per unit time for each data set was then plotted and compared against the PMF of the Poisson distribution with the same mean.

### **Results:**

After performing the statistical analysis it was observed that the arrival rate of the the incoming packages was higher than the rate of outgoing as shown in figure 1.

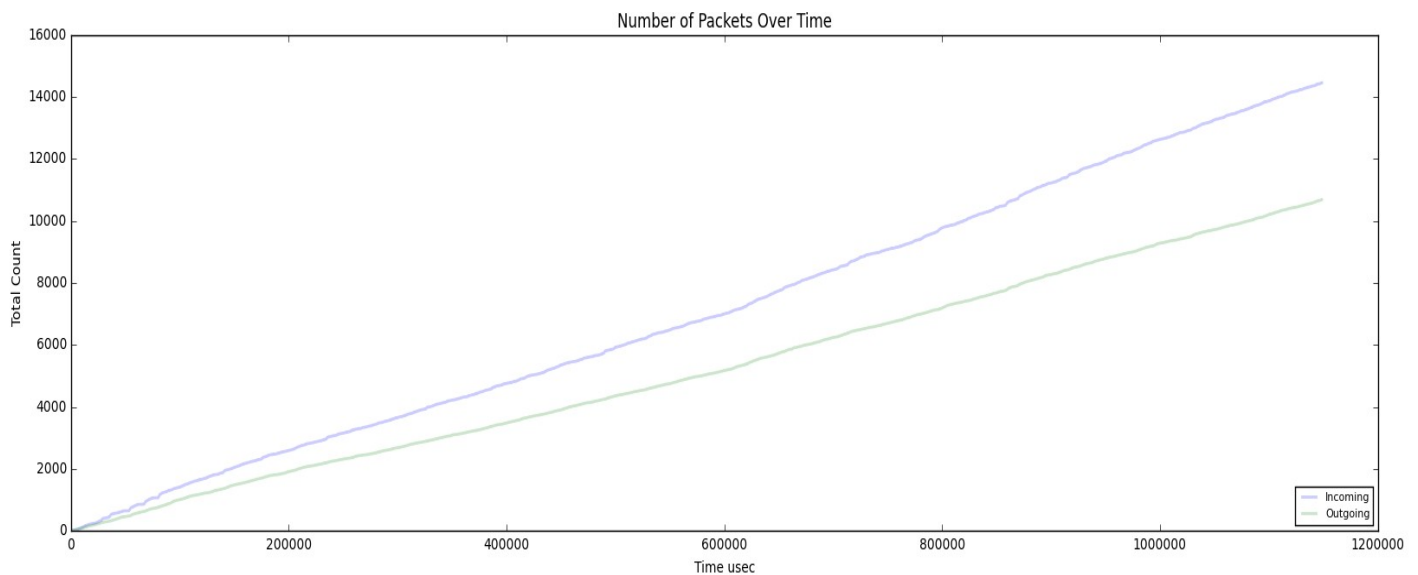


Figure 1. counting Process of packets over time fo bothi incoming and outgoing traffic

The inter-packet time probability distributions appeared exponential concentrating around small sized packets, with the incoming distribution having a much larger spread in size. Upon closer inspection the outgoing inter-packet time distribution might not exactly fit an exponential model as shown in figure 2.

I

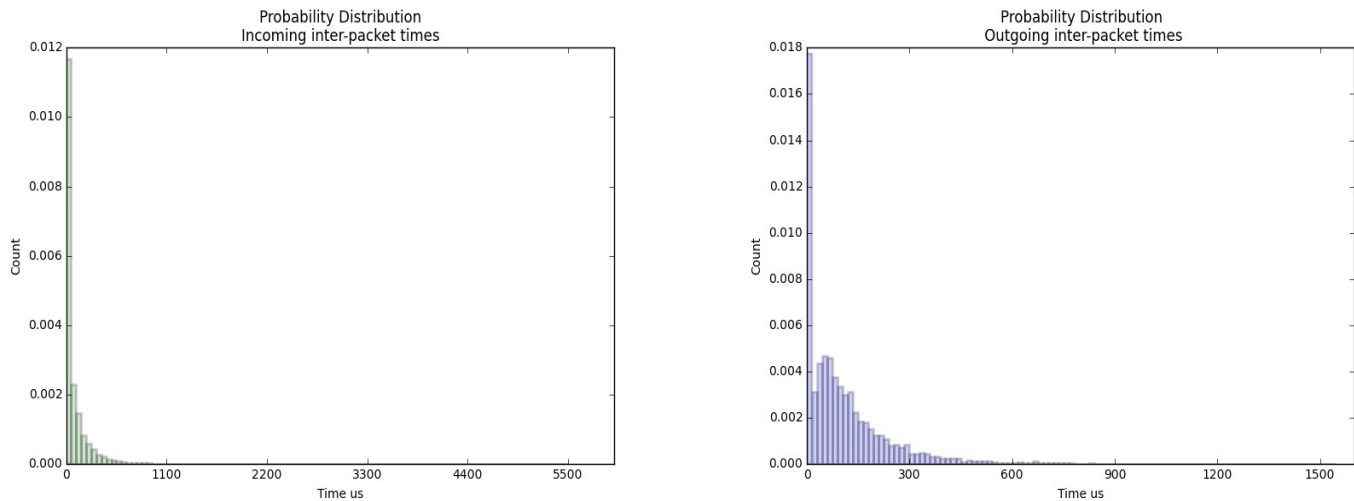


Figure 2. Inter Packet time probability distribution

Finally looking at the arrival count distribution the outgoing arrival count probabilities seemed to most resemble the Poisson distribution with the same mean as shown in figure 3.

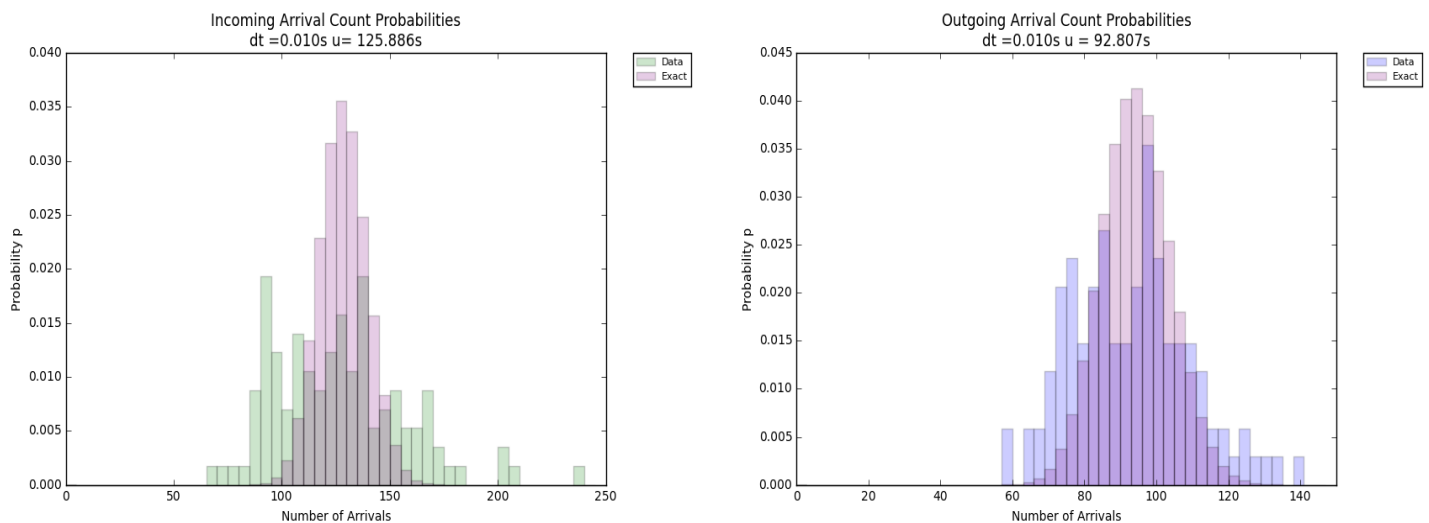


Figure 3. Arrival count distribution vs. Poisson Distribution with the same mean

Using the program the following data was also aquired:

```

////////////////// INCOMING PACKET STATISTICS ////////////////////
N: 14454
Total Bytes: 14541135
Inter-Arrival Mean: 79.456
Inter-Arrival Variance: 21801.989

```

//////////////// OUTGOING PACKET STATISTICS //////////////////

N: 10681

Total Bytes: 4084215

Inter-Arrival Mean: 107.516

Inter-Arrival Variance: 15564.572